Draft Guidance for Industry and FDA Staff

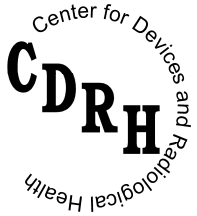
Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.
Document issued on: October 21, 2009

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Alternatively, electronic comments may be submitted to http://www.regulations.gov. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft guidance document contact Nicholas Petrick (OSEL) at 301-796-2563, or by e-mail at Nicholas.Petrick@fda.hhs.gov; or Joyce Whang (ODE) at 301-796-6516, or by e-mail at Joyce.Whang@fda.hhs.gov.



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Imaging and Applied Mathematics
Office of Science and Engineering Laboratories

Radiological Devices Branch
Division of Reproductive, Abdominal, and Radiological Devices
Office of Device Evaluation

Preface

Additional Copies

Additional copies are available from the Internet at: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187249.htm. You may also send an e-mail request to dsmica@fda.hhs.gov to receive an electronic copy of the guidance or send a fax request to 301-847-8149 to receive a hard copy. Please use the document number (1697) to identify the guidance you are requesting.

Table of Contents

Guidance for Industry and FDA Staff

Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

1. Introduction

This draft guidance document provides recommendations to industry, systems and service providers, consultants, FDA staff, and others regarding premarket notification (510(k)) submissions for computer-assisted detection (CADe[1]) devices applied to radiology images and radiology device data (often referred to as "radiological data" in this document). CADe devices are computerized systems that incorporate pattern recognition and data analysis capabilities (i.e., combine values, measurements, or features extracted from the patient radiological data) and are intended to identify, mark, highlight, or in any other manner direct attention to portions of an image, or aspects of radiology device data, that may reveal abnormalities during interpretation of patient radiology images or patient radiology device data by the intended user (i.e., a physician or other health care professional), referred to as the "clinician" in this document. In drafting this document, we considered the recommendations on documentation and performance testing for CADe devices made during the Radiology Advisory Public Panel on March 4-5, 2008.[2] This draft guidance is issued for comment purposes only.

[1] The use of the acronym CADe for computer-assisted detection may not be a generally recognized acronym in the community at large. It is used here to identify the specific type of devices discussed in this document.

[2] http://www.fda.gov/ohrms/dockets/ac/cdrh08.html#radiology

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

The Least Burdensome Approach

This draft guidance document reflects our careful review of what we believe are the relevant issues related to computer-assisted detection on radiological data and what we believe would be the least burdensome way of addressing these issues. If you have comments on whether there is a less burdensome approach, however, please submit your comments as indicated on the cover of this document.

2. Background

This draft guidance applies to the CADe devices identified in **Section 3. Scope** by their classification regulation (21 CFR 892.2050) and product codes (NWE, OEB, OMJ). A manufacturer who intends to market one of these devices must:

- conform to the general controls of the Federal Food, Drug, and Cosmetic Act (the Act), including the premarket notification requirements described in 21 CFR 807 Subpart E;
- conform to the special controls designated for this device (see 21 CFR 892.2050(b)); and
- obtain a substantial equivalence determination from FDA prior to marketing the device. (See also 21 CFR 807.81 and 807.87.)

This document provides recommendations regarding premarket notifications (510(k)s) for these devices. It supplements the requirements in 21 CFR 807.87 and other FDA documents concerning the specific content of a premarket notification submission, including the guidance, **Format for Traditional and Abbreviated 510(k)s**.[3]

Under "**The New 510(k) Paradigm - Alternate Approaches to Demonstrating Substantial Equivalence in Premarket Notifications**,"[4] a manufacturer may submit a Traditional 510(k) or has the option of submitting either an Abbreviated 510(k) or a Special 510(k). FDA believes an Abbreviated 510(k) provides the least burdensome means of demonstrating substantial equivalence for a new device, particularly once FDA has issued a guidance document addressing

[3] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm084365.htm

[4] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm080187.htm

that device. Manufacturers considering certain modifications to their own cleared devices may lessen the regulatory burden by submitting a Special 510(k).

3. Scope

This document provides guidance regarding premarket notification (510(k)) submissions for CADe devices applied to radiology images and radiology device data. Radiological data include those that are produced during patient examination with ultrasound, radiography, magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), etc.[5] As stated above, CADe devices are computerized systems intended to identify, mark, highlight, or in any other manner direct attention to portions of an image, or aspects of radiology device data, that may reveal abnormalities during interpretation of patient radiology images or patient radiology device data by the clinician. This draft guidance covers CADe devices marketed as a complete package with a review workstation, or as an add-on software to be embedded within imaging equipment, an image review platform (for example, a PACS (picture archiving and communications system)), or other imaging accessory equipment.

This draft guidance document applies to the CADe devices under 21 CFR 892.2050 Picture archiving and communications systems, and the following current product codes:

- NWE (Colon computed tomography system, computer-aided detection),
- OEB (Lung computed tomography system, computer-aided detection), and
- OMJ (Chest x-ray, computer-aided detection).

This draft guidance does not address non-CADe device components or capabilities, including the many non-CADe devices that are covered by 21 CFR 892.2050, i.e. product codes LLZ (System, Image Processing, Radiological) and NFJ (System, Image Management, Ophthalmic).

21 CFR 892.2050 Picture archiving and communications system.

(a) *Identification.* A picture archiving and communications system is a device that provides one or more capabilities relating to the acceptance, transfer, display, storage, and digital processing of medical images. Its hardware components may include workstations, digitizers, communications devices, computers, video monitors, magnetic, optical disk, or other digital data storage devices, and hardcopy devices. The software components may provide functions for performing operations related to image manipulation, enhancement, compression or quantification.

---

[5] For any use of a contrast imaging agent, we recommend that you verify that such comports with the regulation, labeling, and indications of the imaging drugs and devices. You may wish to consult the draft guidance **New Contrast Imaging Indication Considerations for Devices and Approved Drug and Biological Products (DRAFT)** (http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126051.pdf) for new contrast imaging drugs and devices indications.

(b) *Classification.* Class II (special controls; voluntary standards--Digital Imaging and Communications in Medicine (DICOM) Std., Joint Photographic Experts Group (JPEG) Std., Society of Motion Picture and Television Engineers (SMPTE) Test Pattern).

By design, a CADe device can be a unique detection scheme specific to only one type of potential abnormality or a combination or bundle of multiple parallel detection schemes, each specifically designed to detect one type of potential abnormality that is revealed in the patient radiological data. Examples of CADe devices that fall within the scope of this draft guidance include:

- a CADe device designed to identify and prompt colonic polyps on CT colonography studies,
- a CADe designed to identify and prompt filling defects on thoracic CT examination, and
- a CADe designed to identify lung nodules on MRI studies.

This draft guidance does not cover devices in the Class III product code MYN (Analyzer, Medical Image), any CADe devices that are intended for use during intra-operative procedures, or any computer-assisted diagnostic devices (CADx) or computer-triage devices, whether marketed as unique devices or bundled with a computer-assisted detection device that, by itself, may be subject to this draft guidance. Below is further explanation of the CADx and computer-triage devices not covered by this draft guidance:

- CADx devices are computerized systems intended to provide information beyond identifying, marking, highlighting, or in any other manner directing attention to portions of an image, or aspects of radiology device data, that may reveal abnormalities during interpretation of patient radiology images or patient radiology device data by the clinician. CADx devices include those devices that are intended to provide an assessment of disease or other conditions in terms of the likelihood of the presence or absence of disease, or are intended to specify disease type (i.e., specific diagnosis or differential diagnosis), severity, stage, or intervention recommended. An example of such a device would be a computer algorithm designed both to identify and prompt lung nodules on CT exams and also to provide a probability score to the clinician for each potential lesion as additional information.

- Computer-triage devices are computerized systems intended to in any way reduce or eliminate any aspect of clinical care currently provided by a clinician, such as a device for which the output indicates that a subset of patients (i.e., one or more patients in the target population) are normal and therefore do not require interpretation of their radiological data by a clinician. An example of this device is a prescreening computer scheme that identifies patients with normal MRI scans that do not require any review or diagnostic interpretation by a clinician.

For any of these types of devices, we recommend that you contact the Agency to inquire about premarket pathways, regulatory requirements, and recommendations about nonclinical and clinical data.

4. Describing the Device in a 510(k) Premarket Notification

We recommend you identify your device by the regulation and product code described in **Section 3. Scope**, and provide an overview of your CADe algorithm and a detailed description of the following:

- the algorithm design and function,
- processing steps,
- features,
- models and classifiers,
- training paradigm,
- databases,
- reference standard, and
- scoring methodology.

General Information

In accordance with 21 CFR 807.87, provide proposed labels, labeling, and advertisements sufficient to describe the device, the intended use, directions for use, a complete description of the operational principles for your device, and a 510(k) summary or a 510(k) statement (see 21 CFR 807.87(e), (f) & (h) and **Section 8. Labeling**). In providing a description of your device, we recommend you include the following information:

- target population information including patient population, organs of interest, diseases/conditions/abnormalities of interest, and appropriate clinician intended to use the device (e.g., radiologist, family practice physician, nurse);
- radiological data used as input and compatible with your CADe design, including imaging modalities (e.g., computed tomography, magnetic resonance), make, model and specific trade name for each modality/system if applicable, specific image acquisition parameter ranges (e.g., kVp range, slice thickness), and specific clinical imaging protocol(s) (e.g., oral contrast studies, magnetic resonance imaging (MRI) sequence);
- current clinical practice relevant to the diseases/conditions/abnormalities of interest;
- proposed clinical workflow (as compared to the predicate device) including a description of:
  - how your device is labeled for use in clinical practice,
  - when your device should be utilized within the proposed workflow,
  - effects on interpretation time as it relates to specific claims;

- device impact (as compared to the predicate device), including:
  - the impact on patient health from additional medical procedures resulting from an unnecessary patient recommendation or follow-up by the clinician based on the information provided by the device (e.g., an incorrect follow-up determination would likely result in short term surveillance imaging for the patient or an incorrect follow-up determination would likely result in a biopsy),
  - the impact on the patient associated with device performance for true positive and true negative marks, separately, and
  - the impact on the patient associated with device performance for false positive and false negative marks, separately;
- device limitations (as compared to the predicate device) including diseases/conditions/abnormalities for which the device has been found ineffective and should not be used; and
- supporting data from the scientific literature.

Algorithm Design and Function

We recommend you provide information on the algorithm design and function including details on the following:

- algorithm implementation:
  - a description of the format of all CADe marks available, including all relevant geometric and other properties such as shape, size, intended location in relation to region of interest (e.g., overlap, adjacent), border (e.g., solid, dashed), and color.

We recommend you provide a detailed flowchart identifying the processing, features, models, and classifiers utilized by your algorithm. We suggest your flowchart include the following:

- all manual operations and associated predefined default settings (e.g., selection of rules or thresholds by the physician);
- all semi-automatic operations and associated predefined default settings (e.g., selection of seed points for region segmentation); and
- all automatic operations that do not involve direct interaction with the clinician.

You should include other algorithm information including:

- name,
- version and important characteristics of the software platform,
- operating system, and
- programming language.

We also recommend you describe the design and function for each stage of your algorithm, where a stage is an independent or well-defined functional unit within the CADe algorithm. Your description may likely include a discussion of the following:

- purpose of the stage,
- processing steps,
- features,
- classifiers and their estimated complexity,
- training paradigm,
- development and training databases utilized, and
- reference standard.

Processing

Processing refers to any image or signal normalization, filtering, and segmentation of areas or structures of interest. Examples of filtering and segmentation processes are the use of a smoothing filter for noise reduction or the delineation of an organ of interest from its surroundings, respectively. We recommend that you provide a description of all processing as well as relevant algorithm flowcharts, equations, and references.

Normalization processing refers to calibration or transformation of image or signal characteristics to that of a reference image or signal. We recommend you provide a description of the technique used to establish the proper calibration or transformation, as well as the characteristics of the reference.

Features and Feature Selection

Features are computer or human estimated quantities characterizing images, regions, or pixels within radiological data, including any specific patient characteristics (e.g., age, sex, ethnicity). Feature selection includes any processes used to cull a set of candidate features. Feature selection or dimensional reduction may be accomplished by manual selection of important features by a user or by an automated selection algorithm (e.g., through the use of a genetic algorithm). For each stage of your algorithm, we recommend you provide:

- the total number of features computed and evaluated during algorithm development, and
- the number of features retained after feature selection, if appropriate.

For each feature, we recommend you provide:

- a description of how the feature is determined (e.g., mathematical expression),
- the feature class (e.g., demographic, biological, morphological and geometrical features), and

- the feature type (i.e., computer estimated feature value or reader estimated feature value).

Models and Classifiers

We define a model as any method or rule used to rate or categorize information within an image. A classifier is a human- or statistically-defined model used to rate or categorize regions within an image with respect to disease, condition, or abnormality. This model is an assumed relationship between image features and the rating or categorization of disease, condition, or abnormality, and depends on a specific set of parameters that are determined in processing steps either manually or automatically. Models and classifiers typically perform some type of pattern recognition procedure. They can vary from a single threshold on a uniquely extracted feature to a complex classifier (i.e., a weighted combination of feature values). For each stage of your algorithm, we recommend you provide the following:

- the number of different models and classifiers utilized; and
- the types of models and classifiers used (e.g., simple threshold, decision tree, linear discriminant, neural network, support vector machine), including specific parameters and values being utilized.

Algorithm Training

Algorithm training is a procedure used to set algorithm parameters and thresholds. This procedure includes the adjustment of filter parameters, the selection of the most discriminant features, and the adjustment of classifier weights and model parameters. Training may be done manually by humans (e.g., the programmer or a medical professional), automatically using a specialized training algorithm, or by a combination of both. For the individual stages as well as the overall algorithm, we recommend you describe your algorithm training paradigm, including the technique employed for feature selection, and indicate if it is performed:

- manually by humans;
- automatically using a computerized training method; or
- by a combination of manual and computerized techniques.

If algorithm training is performed manually, we recommend you provide the number and qualifications of the individuals performing the training. Whether the training is performed manually, automatically, or by a combination of techniques, we recommend you describe the criteria and performance metrics used to determine the settings (i.e., thresholds, weights, or parameters) of each individual stage and provide a summary of the resulting observed performance.

We further recommend you provide history of the accrual and use of data in the training and evaluation of the CADe device.

Databases

Databases refer to the sets of radiology images or radiology device data used in training and testing your device. These databases may contain computer simulated data, phantom data, or patient data depending on the nature of the evaluation.

For a database of computer simulated or phantom data (i.e., training and testing cases), we recommend you provide:

- a description of the phantom or simulation methodology; and
- any data characterizing the relationship between the simulated or phantom data and actual patient data for the imaging technique, organ, and disease of interest.

For each database of patient data (i.e., training and testing cases), we recommend you provide specific information including:

- the patient demographic data (e.g., age, ethnicity, race);
- the patient medical history relevant to the CADe application;
- the patient disease state and indications for the radiologic test;
- the conditions of radiologic testing, for example technique (including whether the test was performed with/without contrast, contrast type and dose per patient, patient body mass index, radiation exposure, T1-weighting for MRI images) and views taken;
- a description of how the imaging data were collected (e.g., make and model of imaging devices and the imaging protocol) and the expertise of the person collecting the data (e.g., x-ray technician);
- the collection sites;
- the processing sites, if applicable (e.g., patient data digitization);
- the number of cases:
  - the number of diseased cases,
  - the number of normal cases,
  - any methods used to determine disease status, location and extent (see **Section 4**, subsection **Reference Standard**);
- the case distributions stratified by relevant confounders or effect modifiers, such as lesion type (e.g., hyperplastic vs. adenomatous colonic polyps), lesion size, lesion location, disease stage, organ characteristics, concomitant diseases, imaging hardware (e.g., makes and models), imaging or scanning protocols, collection sites, and processing sites (if applicable);
- a comparison of the clinical, imaging, and pathologic characteristics of the patient data compared to the target population; and
- a history of the accrual and use of both training and test databases.

For CADe devices intended to be used with proprietary imaging devices, we recommend you provide the trade names, regulatory status, and physical characteristics of these proprietary imaging devices.

Reference Standard

For purposes of this document, the reference standard (also often called the "gold standard" or "ground truth" in the imaging community) for patient data indicates whether or not the disease/condition/abnormality is present and may include such attributes as the extent or location of the disease/condition/abnormality CADe device development and evaluation often relies on databases of a radiology images or radiology device data with a reference standard addressing whether or not the disease/condition/abnormality is present within an individual patient and if so, its location and extent. We refer to this characterization of the reference standard for the patient, e.g., disease status, as the truthing process.

The methodology utilized to establish the reference standard can impact reported performance. The types and nature of the abnormalities marked or not marked by your CADe device should be consistent with the intended use of your device. You should provide the rationale and describe the procedure for defining if a disease/condition/abnormality is present and the location and extent of the disease/condition/abnormality (e.g., based on pathology or based on a standard of care determination). You should also indicate if the reference standard is based on:

- the output from another device;
- an established clinical determination (e.g., biopsy, specific laboratory test);
- a follow-up clinical imaging examination;
- a follow-up medical examination other than imaging; or
- an interpretation by reviewing clinician(s) (i.e., truther(s)).

The methodology utilized to make this reference standard determination should be described and should be fixed prior to initiating your evaluation. For truthing that relies on the interpretation by reviewing clinician (i.e., truther), we recommend you provide:

- the number of truthers involved;
- their qualifications;
- their levels of experience and expertise;
- the instructions conveyed to them prior to participating in the truthing process;
- all available clinical information from the patient utilized by them in the identification of disease/condition/abnormality and in the marking of the location and extent of the disease/condition/abnormality; and
- any specific criteria used as part of the truthing process.

When multiple truthers are involved, you should describe the process by which their interpretations are combined to make an overall reference standard determination and how your process accounts for any inconsistencies between clinicians participating in the truthing process (truth variability). Clinicians participating in the truthing process should not be the same as those who participate in the core clinical performance assessment of the CADe device because doing so would introduce bias into the study results.

Scoring

In addition to determining the reference standard for the location and extent of the disease/condition/abnormality, CADe device development and evaluation often rely on determining whether the spatial location and extent of a CADe mark correspond to the location and extent of the disease/condition/abnormality. We define the procedure for determining the correspondence between the CADe output and the reference standard (e.g., disease location) as the scoring process. The scoring procedure and the scoring definition are important components for interpreting standalone device performance and for appropriately labeling the device.

In this document we describe the scoring used to evaluate device standalone performance. A different type of scoring is used in the clinical performance assessment which is described in the draft guidance entitled **Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions**.[6]

The scoring process should be consistent with the abnormalities being marked by the CADe and the intended use of your device. The scoring process should be described and primary and secondary endpoints should be fixed prior to initiating your evaluation. In your description of the scoring process, we recommend you indicate whether the scoring is based on:

- electronic or non-electronic means;
- physical overlap of the boundary, area, or volume of the mark in relation to the boundary, area, or volume of the reference standard;
- relationship of the centroid of the mark to the boundary or spatial location of the reference standard;
- relationship of the centroid of the reference standard to the boundary or spatial location of the mark;
- interpretation by reviewing readers; or

---

6 http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187277.htm

- other methods.

For scoring that relies on interpretations by reviewing readers, we recommend you provide the number of readers involved, their qualifications, their level of experience and expertise, the specific instructions conveyed to them prior to participating in the scoring process, and any specific criteria used as part of the scoring process. When multiple readers are involved in scoring, you should describe the process by which their interpretations are combined to make an overall scoring determination or how their interpretations are incorporated in the performance evaluation, including how any inconsistencies are addressed.

Other Information

We recommend that you include information for software-controlled devices described in **Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices**[7] and in **Guidance for Off-the-Shelf Software Use in Medical Devices**.[8] The kind of information we recommend is determined by the "level of concern," which is related to the risks associated with a software failure. The level of concern for a device may be minor, moderate, or major. Based on prior CADe device submissions, the level of concern for a CADe system is generally moderate or major.

If the CADe system is an add-on software to be installed within a third party image review platform, we recommend you also provide the names, version/model numbers, and characteristics of these third party platforms as well as a description of the file format of the CADe output that is generated by your device. If applicable, we recommend you refer to **Guidance for the Submission of Premarket Notifications for Medical Image Management Devices**.[9]

We recommend submitting electronically the data used in any statistical analysis in your study. For more information on submitting data electronically, please see the FDA white paper entitled **Clinical Data for Premarket Submissions**.[10]

[7] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089543.htm

[8] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm073778.htm

[9] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm073720.htm

[10] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/ucm136377.htm

5. Standalone Performance Assessment

Because each new CADe device represents a new implementation of software, FDA expects that each new CADe device (as well as software and other design, technology, or performance changes to an already cleared CADe device) will have different technological characteristics from the legally marketed predicate device even while sharing the same intended use. Accordingly, under section 513(i)(1)(A) of the Act, determinations of substantial equivalence will rest on whether the information submitted, including appropriate clinical or scientific data, demonstrate that the new or changed device is as safe and effective as the legally marketed predicate device and does not raise different questions of safety and effectiveness than the predicate device.

To support a substantial equivalence determination for a new CADe device, or for changes to an already cleared CADe device that could significantly affect safety or effectiveness, we recommend you measure and report the performance of your CADe device by itself, in the absence of any interaction with a clinician (i.e., standalone performance assessment). These measurements estimate how well the CADe device, by itself, marks regions of known abnormalities and how well the CADe device avoids marking regions other than the abnormalities (e.g., normal organ and structures). Study endpoints should be selected to establish meaningful and statistically significant performance for the device.

To support substantial equivalence, we recommend comparing the standalone performance of your CADe device to the standalone performance of the predicate device on the same dataset, if possible. Otherwise, the characteristics or makeup of the database used to assess standalone performance should be comparable to the characteristics or makeup of the database used in assessing the predicate device.

The types and nature of the abnormalities marked or not marked by your CADe device should be consistent with the intended use of your device. To measure standalone performance, the true location of abnormalities should be determined through some well-described truthing process (see **Section 4**, subsection **Reference Standard**). The location and extent of a CADe mark should be compared to the truthed location and extent of an abnormality using the established scoring process (see **Section 4**, subsection **Scoring**). The reference standard definition, scoring process, and analysis methodology, including primary and secondary performance endpoints, should be established prior to the collection of the standalone performance assessment data and analysis of these data. Any performance claims based on a covariate analysis should be demonstrated through a prespecified analysis plan.

We recommend that you perform standalone testing in a way that will provide good estimates of performance stratified by important covariates, such as lesion type, size or shape. This stratified standalone performance is useful in labeling by providing the end users with additional information to better interpret the meanings of the CADe marks.

Study Population

We recommend you assess and report your device standalone performance on testing data that is independent and sequestered from the data on which the CADe was developed and trained. Reusing test data (i.e., conducting multiple tests on the same data) is problematic for interpreting the results. Test data, used once before, does not constitute independent data for testing a CADe device because the CADe algorithm may have become trained to that data, either implicitly or explicitly. If you intend to reuse test data, we recommend that you contact the Agency to discuss the scientific validity of your proposed methodology and seek advice on the reuse of test data.

Your testing database should be representative of the target population and the target disease, condition, or abnormality for which your device is intended. We recommend that you provide the protocol for your case collections. An acceptable approach for acquiring data that is representative of the intended use population is the collection of consecutive cases from each participating collection site that fall within the inclusion and outside the exclusion criteria. The full range of diseased/abnormal and normal cases should be sufficiently represented in the testing database.

Enrichment with diseased/abnormal cases is permissible for an efficient and least burdensome representative case dataset but may affect standalone performance estimates (e.g., the performance estimates may not generalize to the intended use population). You may choose to enrich the study population with patient cases that contain imaging findings (or other image data) that are known to challenge clinicians but that still fall within the device's intended use population (i.e., stress testing). For example, if assessing a CADe device designed to detect colon polyps, the study population may be enriched with cases containing smaller polyps. The study should contain a sufficient number of cases from important cohorts (e.g., subsets defined by clinically relevant confounders, effect modifiers, and concomitant diseases) such that standalone performance estimates can be obtained for these individual subsets (e.g., performance estimates for different nodule size categories when evaluating a lung CADe device). Powering these subsets for statistical significance may not be necessary unless specific subset performance claims are being included. A good study design might include and report results for both an enriched data set containing relevant confounders as well as a set of consecutive cases from each participating collection site where the consecutive cases may better represent the standalone performance in clinical practice.

The sample size of the study should be large enough such that the study has adequate power to detect with statistical significance your proposed performance claims. If performance claims are proposed for individual subsets, then the sample sizes for these subsets should be determined accordingly to detect these claims with statistical significance. For formal subset analysis, a pre-specified statistical adjustment for the testing of multiple subsets would be statistically necessary.

As part of the device standalone performance assessment, you should describe the testing database (**see Section 4**, subsection **Databases**). We recommend your performance testing include:

- detection accuracy testing,
- localization accuracy testing,
- reproducibility testing,
- stability analysis, and
- algorithm training performance.

Detection Accuracy

We recommend you estimate and report the CADe standalone performance following the scoring process (see **Section 4**, subsection **Scoring**). The definition of a true positive, true negative, false positive, and false negative CADe mark should be consistent with the intended use of the device. For example, if the device is intended to detect all abnormalities (e.g., benign and malignant), then a true positive CADe mark should be defined as "marking" any abnormalities. On the other hand, if a device is intended to detect only a subset of abnormalities (e.g., only those lesions with certain imaging features), then a true or false CADe mark should be defined accordingly.

For truthing (e.g., disease type, location, and extent) that relies on the interpretation by reviewing readers, we recommend that you account for reader variability in the truthing process and for various consensus or agreement rules between expert readers, in the CADe standalone performance estimates. One method of accounting for variability in the reference standard is to resample the expert truthing panel. See Miller *et* al.[11] for details on one approach.

We recommend you report the overall lesion-based, patient-based, and any other relevant anatomical or image unit-based sensitivities, and average number of false positives per case (FPs/case) or other relevant measure of specificity, at each device operating point as well as stratified analysis per relevant confounder or effect modifier as appropriate (e.g., lesion size, lesion type, imaging or scanning protocols, imaging or data characteristics). FPs/case or other relevant measure of specificity should be derived from normal and abnormal patient data separately. If your device allows the clinician to select or manipulate the device operating point, we recommend you provide the device performance for each selectable operating point or for the range of possible operating points. The detection accuracy assessment methodology, including the selection of primary and secondary performance endpoints, should be determined and fixed prior to initiating your evaluation.

---

[11] Miller, D. P., O'Shaughnessy, K. F., Wood, S. A., and Castellino, R. A., "Gold standards and expert panels: A pulmonary nodule case study with challenges and solutions," Proc. of the **SPIE**, Medical Imaging; 5372: 173–184, 2004.

All performance measures should be reported with associated confidence intervals (CIs). We recommend you provide a description of your methodology for estimating these CIs and the clinical significance associated with these CIs.

We also recommend you provide graphs of the free-response receiver operating characteristic (FROC) curves (i.e., a plot of patient-based sensitivity vs. average number of FPs/case as a function of operating point) when reporting detection accuracy and the clinical interpretation of this analysis. Associated FROC CIs should be reported when appropriate. Resampling techniques, such as bootstrapping,[12] are potential methodologies for estimating these CIs.

Localization Accuracy

Localization accuracy depends upon the scoring criteria used to determine the nature of each CADe detection, i.e., true positive (TP) or false positive (FP). Using only one scoring criterion, i.e., the criterion used for the device performance reported in the labeling (see **Section 4**, subsection **Scoring**), may not be sufficient to evaluate localization accuracy. We recommend you report the CADe localization accuracy by reporting the overall lesion-based, patient-based, and any other relevant anatomical or image unit-based sensitivities, and the average number of FPs/case or other relevant measure of specificity, using multiple scoring criteria. Common scoring criteria used to determine the nature of each CADe detection include:

- centroid of the CADe detection area or volume falling in the reference standard area or volume;
- distance between centroids of the CADe detection and the reference standard;
- ratio of the distance between centroids of the CADe detection and the reference standard, relative to the maximum width of the reference standard region;
- ratio of the area (A) or volume (V) intersection between the CADe detection and the reference standard, with the total area or volume of the reference standard defined as follows:

$$\frac{\mathrm{A}(CAD) \cap \mathrm{A}(Ref)}{\mathrm{A}(Ref)} \quad \text{or} \quad \frac{\mathrm{V}(CAD) \cap \mathrm{V}(Ref)}{\mathrm{V}(Ref)}$$

- ratio of the area (A) or volume (V) intersection between the CADe detection and the reference standard, with the total area or volume of the CADe detection, defined as follows:

$$\frac{\mathrm{A}(CAD) \cap \mathrm{A}(Ref)}{\mathrm{A}(CAD)} \quad \text{or} \quad \frac{\mathrm{V}(CAD) \cap \mathrm{V}(Ref)}{\mathrm{V}(CAD)}$$

[12] Efron, B., and Tibshirani, R., "Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy," **Statistical Science** 1, 54–77, 1986.

- ratio of the area (A) or volume (V) intersection between the CADe detection and the reference standard with the total area or volume union of the reference standard and the CADe detection, defined as follows:

$$\frac{\mathrm{A}(CAD) \cap \mathrm{A}(Ref)}{\mathrm{A}(CAD) \cup \mathrm{A}(Ref)} \quad \text{or} \quad \frac{\mathrm{V}(CAD) \cap \mathrm{V}(Ref)}{\mathrm{V}(CAD) \cup \mathrm{V}(Ref)}$$

We recommend you estimate and report location accuracy performance of your device using various values of the distance and ratio criteria and, if applicable, plots showing the performance change as a function of overlap criteria. The location accuracy assessment methodology, including the selection of primary and secondary performance endpoints, should be determined and fixed prior to initiating your evaluation.

We also recommend you supplement this evaluation by examining the impact of relevant confounders or effect modifiers, such as:

- lesion size,
- lesion type,
- lesion location,
- disease stage,
- organ characteristics,
- imaging hardware,
- imaging or scanning protocol, and
- image or data characteristics (e.g., characteristics associated with differences in digitization architectures for a CADe using scanned films).

We recommend you report all performance measures with associated CIs.

Reproducibility Testing

We recommend you report device reproducibility testing. These testing processes provide insight into the stability of the algorithm and its dependency on parameters usually related to the image acquisition protocol. For example, for digitized image data, the placement of the film in the scanner or the time when the scanning occurs could produce data differences that may affect how the algorithm performs. Providing standalone performance from the same patient and from multiple scans acquired using the same (or a different) acquisition protocol will provide information regarding the reproducibility and stability of the algorithm, with respect to the expected variation in data collection methods. We recommend you provide the following:

- description of the reproducibility study;
- parameters expected to introduce variability in the results (e.g., scanning characteristics, make and model of the imaging devices, acquisition protocol parameters such as contrast agent or probe positioning);

- effects to be monitored (e.g., effect on the segmentation accuracy, feature extraction, overall CADe performance accuracy); and
- results and statistical analysis.

Algorithm Stability Testing

We further recommend you conduct algorithm stability testing including:

- algorithm stability with respect to training set changes (i.e., invariance of the CADe algorithm with respect to the datasets used in its design and training) (e.g., see Yousef *et* al.[13]),
- algorithm stability over time (e.g., invariance to changes in the imaging system, acquisition conditions, operator settings), and
- algorithm stability with respect to other relevant covariates.

For assessment of the stability of your CADe algorithm, we recommend that you describe your methodology and provide results. Such evaluation may be performed, for example, by resampling using multiple bootstrap sets of the training database.

Algorithm Training Performance

We recommend you measure and report standalone performance of your CADe device on the dataset used to train the algorithm. Assessment of the algorithm training performance may include measures such as lesion-based, patient-based, and other relevant anatomical or image unit-based sensitivities, and the average number of false positives per case (FPs/case) or other relevant measure of specificity, at each device operating point. If your device allows clinicians to select or manipulate the device's operating point, we recommend you provide the device performance for individual selectable operating points or the range in performance for continuously varying parameters.

Other Information

In addition to all device performance assessment testing described above, we reiterate our recommendation that you provide a comparison of the performance testing results to the corresponding performances testing results of the legally marketed predicate device to which you are claiming substantial equivalence (e.g., a previously released version of the device), if applicable. Valid comparison of device performance is dependent upon sound study design in the collection of your testing database. We recommend that you describe your comparison analysis, hypothesis to be tested, sample size estimation, and endpoints, and that you provide

[13] Yousef, W. A., Wagner, R. F., and Loew, M. H., "Estimating the uncertainty in the estimated mean area under the ROC curve of a classifier," **Pattern Recognition Letters,** 2005 (http://www.sciencedirect.com/science/article/B6V15-4GTW8JJ-1/2/58c02b75531e668fbcbcd7810c7034b7).

comparison results. For example, when using a common database sequestered from the development and training of both your device and the predicate device, a comparison of the CADe standalone performance may include a measure of the:

- difference in area under the FROC curves with associated statistical analysis (e.g., see Samuelson *et* al.[14]), and
- difference in detection sensitivity and number of FPs/case at the device operating points.

Reporting of standalone performance results may be guided by the FDA Guidance entitled **Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests; Guidance for Industry and FDA Reviewers**.[15]

We again recommend submitting electronically[16] the data used in any statistical analysis in your study including patient information, disease or normal status, lesion size, lesion type, imaging and scanning setting, and imaging and data characteristics.

We also recommend you provide all data on a CD-ROM.

6. Clinical Performance Assessment

As described above, because each new CADe device represents a new implementation of software, FDA expects that each new CADe device (as well as software and other design, technology, or performance changes to an already cleared CADe device) will have different technological characteristics from the legally marketed predicate device even while sharing the same intended use. Accordingly, under section 513(i)(1)(A) of the Act, determinations of substantial equivalence will rest on whether the information submitted, including appropriate clinical or scientific data, demonstrate that the new or changed device is as safe and effective as the legally marketed predicate device and does not raise different questions of safety and effectiveness than the predicate device.

Because the reader is an integral part of the diagnostic process for CADe devices, we believe that a standalone performance assessment without a clinical performance assessment (i.e., a reader study) will usually not be adequate to demonstrate that the diagnostic performance of the CADe device is as safe and effective as the legally marketed predicate. Therefore, you should assume that a clinical assessment will be necessary to demonstrate substantial equivalence between your

---

[14] Samuelson, F. W., and Petrick, N., "Comparing image detection algorithms using resampling," in Proceedings of the IEEE International Symposium on Biomedical Imaging. **IEEE**, pp. 1312–1315, 2006.

[15] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm

[16] See footnote 10.

CADe device and its predicate for its intended use, when used by the intended user and in accordance with its proposed labeling and instructions. This clinical performance assessment should provide an estimate of the clinical effect of the CADe device on clinician performance. If you believe a clinical assessment may not be necessary for demonstrating substantial equivalence of your device with the predicate, we recommend that you contact the Agency to seek advice prior to conducting your studies.

For clinical assessment, various control arms can be employed, including reading aided by the predicate device and unaided reading. The use of the predicate device as the control, with both devices evaluated on the same data set, allows for direct comparison of your device with the predicate for assessing substantial equivalence. The use of unaided reading as the control provides an assessment of the clinical effectiveness of your device, which, in 510(k) studies, should be compared with the clinical effectiveness of the predicate device, as estimated in a prior study. For this comparison to be unbiased, the two studies would ordinarily have to be calibrated on the distributions of important covariates, which can require that the data be available at the patient level in both studies. In addition, the comparison can be problematic to make if different sets of readers, different reference standards, or different scoring methods are used in the two studies.

For further detail on potential clinical assessment methodologies, we recommend that you consult the draft guidance entitled **Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions**.[17]

Examples of changes to an already cleared CADe device for which we recommend submitting a clinical performance assessment include:

- characteristics or makeup of the database used to assess standalone performance (see **Section 5**) cannot be demonstrated to be comparable to the characteristics or makeup of the database used in assessing the predicate device and these difference raises clinical concerns (i.e., could significantly affect safety or effectiveness);
- the results of the standalone performance assessment (see **Section 5**) are different from those of the predicate device, and the significance and effect on the clinician or patient for these different levels of performance are not well-known or well-described in the literature;
- the reference standard definition, scoring process, analysis methodology, or performance endpoints are different from those of the predicate device, and the significance and effect on the clinician or patient of these differences are not well-known or well-described in the literature;

---

17 http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187277.htm

- the algorithm design is different from that of the predicate device and this difference raises clinical concerns (i.e., could significantly affect safety or effectiveness);
- the device design has different human factors from those of the predicate device (e.g., clinician's interaction with a different CADe output display); or
- a new precursor technology or acquisition protocol is employed, changing the nature of the inputs to the CADe (e.g., the current CADe device is applied to digital radiographs whereas the predicate device was applied to film-based radiographs).

There may be situations where a standalone performance assessment without a clinical performance assessment (i.e., a reader study) may be sufficient to demonstrate substantial equivalence. If you believe that a standalone performance assessment without a clinical performance assessment (i.e., a reader study) may suffice to show substantial equivalence, we recommend you contact the Agency to discuss your proposed approach.

7. User Training

We recommend you provide a summary of the procedure that will be used to train the intended users of your device when marketed. The goal of this training should be to help clinicians use the CADe device in an appropriate manner and to provide training so that they can achieve the expected device effectiveness. Training should include both the expected advantages and known limitations of the device (e.g., the CADe does not identify calcified nodules). An aspect of the training may be provided in the form of a self-test for the clinician. This self-test should provide feedback to the clinician on how well he/she performs before and after the integration of the CADe device and guidance on how to improve his/her performance. Training should be based on a broad set of patient data including normal cases. This training data should include typical true positives (TPs) and false positive (FPs) that the device tends to output, as well as typical true negatives (TNs) and false negatives (FNs).

For CADe devices allowing multiple thresholds or operating points, the training should help clinicians identify the most appropriate device setting for their practices. In addition, the training should help allow clinicians to identify suitable CADe reading scenarios.

8. Labeling

The premarket notification must include labeling in sufficient detail to satisfy the requirements of 21 CFR 807.87(e). The following suggestions are aimed at assisting you in preparing labeling that satisfies the requirements of 21 CFR Part 801.[18]

[18] Although final labeling is not required for 510(k) clearance, final labeling must comply with the requirements of 21 CFR Part 801 before a medical device is introduced or delivered for introduction into interstate commerce. In addition, final labeling for prescription medical devices

Your user manual should include the information described below.

Indications for use

We recommend that the indications for use (IFU) address how the device will be used, for example:

> *The device is intended to assist [target users] in their review of [patient/data characteristics] in the detection of [target disease/condition/abnormality] using [image type/technique and conditions of imaging]*.

Directions for use

There must be adequate directions for use as described in 21 CFR 801.5; the requirements applicable to prescription devices are described in 21 CFR 801.109. You should submit clear and concise instructions that delineate the technological features of the specific device and how the device is to be used on patient images/data. Instructions should encourage local/institutional training programs designed to familiarize clinicians with the features of the device and how to use it in a safe and effective manner. The direction should also clearly define the intended user of the device.

Warnings

The warnings should address limitations of the device. For example:

> [*target user*] should not rely solely on the output identified by [*device trade name*], but should perform a full systematic review and interpretation of the entire patient dataset.

Another example may be:

> This CADe device has been found to be ineffective for patients with [*disease/condition/ abnormality*]. This CADe should not be utilized with patients presenting with this [*disease/condition/abnormality*].

Precautions

The precautions should discuss the potential for adverse events associated with the use of the device and recommend mitigation measures. The adverse event discussion should at least include a discussion of potential adverse events associated with an increased workup rate (i.e., events from false-positives) and missed disease/condition/abnormality.

---

must comply with 21 CFR 801.109. Labeling recommendations in this guidance are consistent with the requirements of 21 CFR Part 801.

Device Description

We recommend you include the following in your device description:

- an overview of the algorithm design and features,
- an overview of the training paradigm and the training or development database, and
- a description of the reference standard used for patient data utilized in the development and adjustment of the algorithm.

Clinical Performance Assessment

When appropriate, we recommend you include a summary of the clinical performance assessment including:

- study objectives,
- study design,
- patient population, e.g., age, ethnicity, race,
- number of clinicians and their qualification,
- description of the methodology used in gathering clinical information,
- description of the statistical methods used to analyze the data, and
- study results.

Additional information on reporting clinical performance results can be found in the draft guidance entitled **Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions**.[19]

Standalone Performance Assessment

We recommend you provide a summary of the device standalone performance and reproducibility testing including:

- the scoring criteria used to determine the nature of each region marked by your CADe device;

[19] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187277.htm

- the overall lesion-based, patient-based, and any other relevant anatomical or image unit-based sensitivities, and the average number of FPs/case or other relevant measure of specificity, at each available device operating point;
- the stratified analysis per lesion size, per lesion type, per imaging or scanning protocols, per imaging or data characteristics, as appropriate;
- the confidence intervals (CIs) on each measure; and
- the free-response receiver operating characteristic (FROC) performance, as appropriate.